Modelling heterogeneity of survival in band-recovery data using mixtures

SHIRLEY PLEDGER[1] **& CARL J. SCHWARZ**[2], [1]*Victoria University of Wellington, New Zealand and* [2] *Simon Fraser University, Burnaby, Canada*

ABSTRACT *Finite mixture methods are applied to bird band-recovery studies to allow for heterogeneity of survival. Birds are assumed to belong to one of finitely many groups, each of which has its own survival rate (or set of survival rates varying by time and/or age). The group to which a specific animal belongs is not known, so its survival probability is a random variable from a finite mixture. Heterogeneity is thus modelled as a latent effect. This gives a wide selection of likelihood-based models, which may be compared using likelihood ratio tests. These models are discussed with reference to real and simulated data, and compared with previous models.*

1 Introduction

Individual birds in band-recovery studies frequently have intrinsic differences of survival probability. This may lead to bias in survival estimates, and underestimation of standard errors (Pollock & Raveling, 1982; Nichols *et al.*, 1982; Brownie *et al.*, 1985). The presence of heterogeneity of survival in a data set is characterized by an increasing average survival rate with increasing years since banding, as those remaining in the cohort are mainly the birds with higher survival rates. This is the opposite effect to that of senescence, in which aging birds move to lower survival rates through time, giving a decline in the average survival rate for the cohort.

Burnham & Rexstad (1993) proposed models that included beta distributional variation in individual survival rate. This is an infinite mixture, as the survival parameter is from a continuous distribution. This provided a likelihood framework for testing and comparing the heterogeneous models with the homogeneous models of Brownie *et al.* (1985).

This paper provides an alternative model using a finite mixture, employing

Correspondence: S. Pledger, School of Mathematical and Computing Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand. E-mail: shirley.pledger@vuw.ac.nz

methods similar to those of Norris & Pollock (1995, 1996) and Pledger (2000) for modelling capture probabilities in closed populations. We assume there are G groups of birds, with the intrinsic survival for birds within each group being constant. The finite mixture approach also provides likelihood methods for estimation and model comparison, and a wide range of models is available for comparison and choice.

In Section 2 we give notation and select a model formulation. Section 3 has a detailed discussion of one model with a finite mixture used to model heterogeneity of survival, and in Section 4 this is generalized to show a range of models possible by this method.

Section 5 has a substantial example with many models fitted, and in Section 6 we appraise and compare models using simulations. Section 7 has discussion and conclusions.

2 Model choice and notation

There are two formulations for the band-recovery models:

(1) The (S, f) model used by Brownie *et al.* (1985) and Burnham & Rexstad (1993). S_j is the survival rate in year j. The recovery rate f_j in year j is the product $(1 - S_j)c_j\lambda_j$ of the probabilities for mortality $1 - S_j$, retrieval c_j and band reporting λ_j.
(2) The (S, r) model used by (for example) Catchpole *et al.* (1995) replaces f_j with $(1 - S_j)r_j$, where r_j is a (retrieval and) reporting rate.

Since we may wish to model heterogeneity in S alone, we use the second formulation, which has separated the survival and reporting terms.

The model with (for example) both S and r dependent on time (year), Model 1 of Brownie *et al.* (1985), is specified here as $\{r_t, S_t\}$, in accord with the notation in Lebreton *et al.* (1992). The probability that a bird ringed in year i $(i = 1, 2, \ldots k)$ dies and has its band reported in year j $(i \leqslant j \leqslant l)$ is

$$\text{Prob}(i, j) = S_i S_{i+1} \ldots S_{j-1}(1 - S_j)r_j$$

with the convention that if $i = j$ the S terms form an empty product with value 1, so that

$$\text{Prob}(i, i) = (1 - S_i)r_i$$

The probability of the band not being reported in any of the l reporting years is

$$\text{Prob}(i, l + 1) = 1 - \sum_{y = i}^{l} \text{Prob}(i, y)$$

With this notation, we now consider in detail a model that allows for heterogeneity of survival.

3 Heterogeneity of survival

Heterogeneity between individuals may be allowed in a model by assuming each bird has an intrinsic survival probability, which is a realization of a random variable S. Burnham & Rexstad (1993) used the beta distribution for this part of their modelling. We use instead a distribution with a finite support: each individual has

survival rate S_g with probability $\pi_g, g = 1, 2 \ldots G$, $\Sigma \pi_g = 1$. The use of G groups with constant survival probability within each group is an artefact to introduce heterogeneity: we do not believe there really are G groups. Mixture models often have little power to discriminate between finite and infinite mixtures, or to find the number of groups in a data set (Lindsay, 1995, Section 1.4.2, but see also Richardson & Green, 1997). Models using finite mixtures with two, three or more groups, or the infinite mixture of the beta distribution, frequently give very similar fits to the data.

We now define a model that allows for time effects in the reporting rate and heterogeneity in the survival rate, Model $\{r_t, S_{h_G}\}$. This is a generalization of Model $\{r_t, S_c\}$ with constant S (the (S, r) formulation of Model 2 of Brownie *et al.*, 1985).

Each bird banded in year i has probability π_g of belonging in group $g (g = 1, 2 \ldots G)$, in which case

$$\text{Prob}(i, j | \text{Group} g) = S_g^{j-1}(1 - S_g)r_j$$

Thus, the overall probability of its band being reported in year j is

$$\text{Prob}(i, j) | \text{Group} g) = \sum_{g=1}^{G} \pi_g S_g^{j-i}(1 - S_g)r_j$$

As before, if $i = j$ the S terms disappear, and

$$\text{Prob(no recovery of band)} = 1 - \text{total probability of recovery}$$

The model may be fitted by maximizing the log of this likelihood:

$$L = \prod_{i=1}^{k} \left\{ \prod_{j=i}^{l} \left[\sum_{g=1}^{G} \pi_g S_g^{j-i}(1 - S_g)r_j \right]^{x_{ij}} \times \left[1 - \sum_{j=i}^{l} \sum_{g=1}^{G} \pi_g S_g^{j-i}(1 - S_g)r_j \right]^{x_{i,(l+1)}} \right\}$$

where $x_{i,j}$ is the number of bands from the year i cohort reported in year j, and $x_{i,(l+1)}$ is the number of bands not reported. This likelihood provides maximum likelihood estimates and the machinery of goodness-of-fit tests and likelihood ratio tests.

For example, we may compare a two-group complete model $\{r_t, S_{h_2}\}$ with the one-group reduced model $\{r_t, S_c\}$. This is essentially a test for the presence of heterogeneity. The reduced model is at the boundary of the parameter space of the complete model (e.g. at $\pi_1 = 0$), which means the regularity conditions required for a standard likelihood-ratio test are not met. Instead, we may do a non-standard test, using the usual test statistic

$$2(\log L_C - \log L_R)$$

where C and R refer to the complete and reduced models respectively, but testing it against a distribution that is a $50:50$ mixture of zeros and χ_1^2 (Self & Liang, 1987; Pledger, 2000). The significance level for this test is obtained by halving the χ_1^2 p-value. This approximate result is from Self & Liang's Theorem 3 Case 5, although their conditions are not exactly met. They allow for testing if a parameter is on the boundary of the parameter space. In our test, we may take $\pi_1 \leqslant 0.5$ to uniquely label group 1 as the smaller group. When we test H_0: $\pi_1 = 0$, say, if H_0 is true, the parameter S_1 vanishes. This complication is discussed in Davies (1977, 1987). It does not adversely affect the Self & Liang result, and the $50:50$ mixture of zeros and χ_1^2 is still seen to hold by simulation checks (Pledger, 2000). Other

authors have also found this distribution by different methods. This approximate test is quicker than the alternative of using a bootstrap procedure.

We may also test whether more than two groups are needed to adequately represent the data. The extra parameters give different ways of reducing from (say) three groups to two ($\pi_1 = 0$, $S_1 = S_2$, $S_1 = S_3$ or $S_2 = S_2$), and this is not covered in Self & Liang (1987). Simulation confirms that the $50 : 50$ mixture of 0 and χ_1^2 is no longer valid. Here, a bootstrap procedure is recommended.

If the number of groups is regarded as a parameter to be estimated, there is a global maximum likelihood achieved at a certain number of groups, and there is no change in the maximized likelihood if more groups are added (Norris & Pollock, 1996; Lindsay & Roeder, 1992). We may choose the number of groups at which the global maximum likelihood is achieved, or we may fit separate models with increasing numbers of groups, selecting the model beyond which there is no significant increase of maximum likelihood, using the non-standard likelihood ratio test described above. In practice, the two approaches often lead to the same choice of number of groups, and in many cases only two groups are required to provide a parsimonious model that fits the data well.

The introduction of heterogeneity by modelling a parameter as a random variable from a finite mixture may now be used for a variety of new models.

4 More general models

One more general model allows for both time (year) and heterogeneity effects in survival. If the bird groups have survival rates following similar patterns through time, although at different levels, the time and heterogeneity effects may be made additive on the log scale (e.g. the proportional effect model of Catchpole *et al.*, 1995). However, we use the alternative of linearity on the logistic scale, not least because it maintains survival estimates between 0 and 1. In many situations these alternatives give similar results.

The main effects model, $\{r_t, S_{t+h_G}\}$, has

$$\text{Prob}(i,j) = \sum_{g=1}^{G} \pi_g \left(\prod_{b=i}^{j-1} S_{bg} \right) (1 - S_{jg}) r_j \tag{1}$$

where the survival rate for group g in year j is modelled as

$$\log\left(\frac{S_{jg}}{1 - S_{jg}} \right) = \mu + \tau_j + \eta_g \tag{2}$$

A corner-point parameterization with $\tau_1 = \eta_1 = 0$ makes μ the logit of survival for group 1 at year 1, τ_2 to τ_l adjustments for later years, and η_2 to η_G adjustments for the other groups. The time effect τ_j is fixed, while the heterogeneity effect η_g is a latent effect because each bird's group is unknown. This effect behaves like a random effect in experimental design, with g taking values $1, \ldots, G$ with probabilities $\pi_1, \ldots, \pi_G$ respectively. Equations (1) and (2) are combined with the likelihood

$$L = \prod_{i=1}^{k} \left\{ \prod_{j=i}^{l} [\text{Prob}(i,j)]^{x_{ij}} \times \left[1 - \sum_{j=1}^{l} \text{Prob}(i,j) \right]^{x_{i,(l+1)}} \right\}$$

to give the model formulation.

The special case with all $\tau_j = 0$ reduces to a model equivalent to $\{r_t, S_{h_G}\}$. Hence, a test of H_0: all $\tau_j = 0$ is testing if year has any effect on survival rate, while allowing for heterogeneity. This is a standard likelihood ratio test with χ^2 on $l - 1$ degrees of freedom (Self & Liang, 1987). Similarly, if $G = 2$, a test of H_0: $\eta_2 = 0$ tests for the presence of heterogeneity, while allowing for time effects. This is another non-standard likelihood ratio test, tested on a $50:50$ mixture of $0 : \chi_1^2$.

It is possible to model survival with interactive time and heterogeneity effects. The model $\{r_t, S_{t \times h_G}\}$ assumes

$$\log\left(\frac{S_{jg}}{1 - S_{jg}}\right) = \mu + \tau_j + \eta_g + (\tau\eta)_{jg}$$

with suitable constraints to give $(l - 1)(G - 1)$ independent interaction parameters $(\tau\eta)_{jg}$ (e.g. $(\tau\eta)_{j1} = (\tau\eta)_{1g} = 0$ for all j and g). This model allows different patterns over the years for different groups—a year that is good for survival of one group may be poor for another. Hypothesis testing would start with the interaction term, to preserve hierarchical models that do not admit an interaction term without its associated main effects.

If there may be some effect due to years since banding, such as a lower survival rate in the first year after banding, it is allowed for by modelling the survival of group g in year j for year a since banding, S_{jag}, as

$$\log\left(\frac{S_{jag}}{1 - S_{jag}}\right) = \mu + \tau_j + \alpha_a + \eta_g$$

This is model $\{r_t, S_{t+a+h_g}\}$, with α_a providing the (fixed) effect of year since banding (age in the study). We could let $a = 1$ for the first year, and $a = 2$ for subsequent years, and use the constraint $\alpha_1 = 0$. The regular likelihood ratio test of H_0: $\alpha_2 = 0$ would check if there is a different survival rate in the first year, while allowing for time and heterogeneity effects.

The heterogeneity may be put in the reporting rate parameter instead, if that seems a useful model to try. A model with the main effects of time and heterogeneity effects in r and time effects only in S is $\{r_{t+h_G}, S_t\}$, with

$$\text{Prob}(i,j) = \left(\prod_{b=i}^{j-1} S_b\right)(1 - S_j)\sum_{g=1}^{G} \pi_g r_{jg}$$

The reporting rate r_{jg} satisfies

$$\log\left(\frac{r_{jg}}{1 - r_{jg}}\right) = \mu' + \tau_j' + \eta_g' \tag{3}$$

with fixed time effects τ_j' and latent heterogeneity effects η_g'. The prime denotes parameters for r rather than S.

If the possible heterogeneity is thought to extend over both survival and reporting rate, the model would use

$$\text{Prob}(i,j) = \sum_{g=1}^{G} \pi_g\left[\left(\prod_{b=i}^{j-1} S_{bg}\right)(1 - S_{jg})r_{jg}\right]$$

with S_{jg} and r_{jg} modelled as in equations (2) and (3).

All these models, and more, may be adapted for the S, f formulation, if desired. However, this parameterization may not lead to sensible models. For example, if survival rates are modelled using a mixture, then recovery rates (f) are unlikely to be homogeneous as this parameter implicitly includes $(1 - S)$.

There is some parameter redundancy when the interactive models are invoked, and this must be allowed for in counting parameters and finding degrees of freedom.

These models are all fitted by maximum likelihood, which makes available parameter estimates, their standard errors and their profile likelihood intervals.

5 Example

To illustrate these models, we use the San Luis adult male mallard data of Brownie *et al.* (1985, p. 28), with nine banding years and recovery years. The recovery matrix is in Table 1.

Table 2 shows the models fitted, their numbers of independent parameters, residual deviances ($-2 \times$ maximized likelihood), and their Pearson goodness-of-fit χ^2 statistics, degrees of freedom and p-values.

The Pearson goodness of fit tests show that most of the models fit the data adequately. Only models with no allowance for time effects in either r or S fail to give a good fit at the 5% significance level. Similar results are found with the deviance goodness of fit test. The variance inflation factors (Pearson's χ^2/d.f.) have mean 1.146, minimum 0.638 and maximum 1.597, values near enough to 1 to indicate no serious overdispersion problem.

For model comparison, we first consider the four models with constant parameters or time effects only, and their counterparts with heterogeneity of survival added. In each case the model was improved by modelling two bird groups for heterogeneity. The four tests are all non-standard ones, on a $50:50$ mixture of $0:\chi_1^2$.

$\{r_c, S_c\}$ versus $\{r_c, S_{h_2}\}$: test statistic 6.647, $p = 0.0014$
$\{r_t, S_c\}$ versus $\{r_t, S_{h_2}\}$: test statistic 6.603, $p = 0.0051$
$\{r_c, S_t\}$ versus $\{r_c, S_{t+h_2}\}$: test statistic 3.891, $p = 0.0243$
$\{r_t, S_t\}$ versus $\{r_t, S_{t+h_2}\}$: test statistic 11.871, $p = 0.0003$

Going to three bird groups gives no reduction in residual deviance, indicating that two groups are adequate to represent the heterogeneity.

There was little to choose among the models $\{r_t, S_{h_2}\}$, $\{r_c, S_{t+h_2}\}$ and $\{r_t, S_{t+h_2}\}$.

TABLE 1. Recovery matrix for San Luis adult male mallards

Banding year	Recovery year									Number not recovered
	1	2	3	4	5	6	7	8	9	
1	10	13	6	1	1	3	1	2	0	194
2	0	58	21	16	15	13	6	1	1	518
3	0	0	54	39	23	18	11	10	6	724
4	0	0	0	44	21	22	9	9	3	482
5	0	0	0	0	55	39	23	11	12	803
6	0	0	0	0	0	66	46	29	18	918
7	0	0	0	0	0	0	101	59	30	1060
8	0	0	0	0	0	0	0	97	22	819
9	0	0	0	0	0	0	0	0	21	291

TABLE 2. Models fitted to San Luis adult male mallard data (Brownie *et al.*, 1985, p. 28)

Model	Residual deviance	No. indep. parameters	Pearson χ^2	Goodness of Fit	
				d.f.	p-value
$\{r_c, S_c\}$	8673.916	2	54.3	34	0.0150
$\{r_t, S_c\}$	8650.582	10	32.2	26	0.1865
$\{r_c, S_t\}$	8648.625	10	31.2	26	0.2197
$\{r_t, S_t\}$	8640.555	17	23.6	19	0.2114
$\{r_c, S_a\}$	8667.313	3	47.8	33	0.0457
$\{r_t, S_a\}$	8643.979	11	25.5	25	0.4347
$\{r_c, S_{h_2}\}$	8667.269	4	47.7	32	0.0363
$\{r_t, S_{h_2}\}$	8643.979	12	25.5	24	0.3790
$\{r_c, S_{h.\beta}\}$	8691.203	3	59.3	33	0.0033
$\{r_t, S_{h.\beta}\}$	8646.094	11	27.6	25	0.3266
$\{r_c, S_{t+a}\}$	8654.714	11	37.0	25	0.0577
$\{r_t, S_{t+a}\}$	8632.179	19	14.7	17	0.6158
$\{r_c, S_{t+h_2}\}$	8644.734	12	32.0	24	0.1262
$\{r_t, S_{t+h_2}\}$	8628.684	20	13.9	16	0.6058
$\{r_c, S_{a+h_2}\}$	8667.270	5	47.9	31	0.0268
$\{r_t, S_{a+h_2}\}$	8643.978	13	25.5	23	0.3249
$\{r_c, S_{t+a+h_2}\}$	8644.077	13	26.1	23	0.2933
$\{r_t, S_{t+a+h_2}\}$	8626.230	21	9.6	15	0.8458
$\{r_{t+a}, S_c\}$	8644.358	11	25.9	25	0.4110
$\{r_{t+a}, S_t\}$	8636.595	19	19.2	17	0.3157
$\{r_{t\times a}, S_c\}$	8631.661	18	12.7	18	0.8078
$\{r_{t\times a}, S_t\}$	8624.326	26	8.0	10	0.6298

The first versus the third has $p = 0.0557$, and the second versus third has $p = 0.0417$. All three have a good fit to the data, as seen in the goodness of fit tests in Table 2. The infinite mixture model with individual survival rates based on a beta distribution was also tried. This is the (S, r) version of the (S, f) model in Burnham & Rexstad (1993). This beta model, $\{r_c, S_{h.\beta}\}$, was also in the same group of 'best' models, as indicated by high p-values in Table 2. We did not use the Akaike information criterion (AIC) for model comparison, as the standard AIC does not apply to mixture models.

These three models exhibit similar and interesting behaviour. We consider Model $\{r_t, S_{h_2}\}$ in detail. It distinguishes a small group (8%, $\hat{\pi}_1 = 0.08$) of birds with very low survival rate ($\hat{S}_1 = 0.0064$), and the rest (92%) with survival rate $\hat{S}_2 = 0.6783$. (This gives an average survival estimate of $\pi_1 S_1 + \pi_2 S_2 = 0.6247$, close to $\hat{S} = 0.6375$ from $\{r_t, S_c\}$.) It seems that the recovery matrix is using the opportunity offered by two bird groups to signal one group which dies in the first year after banding and another which progresses through the years with more moderate losses. This led us to try another model for comparison, model $\{r_t, S_a\}$, which reverts to a single group of birds, but introduces an a effect (for 'age in the study', years since banding) allowing for one first-year survival rate and another rate for subsequent years (reminiscent of models for birds ringed as nestlings). This has an almost identical fit to Model $\{r_t, S_{h_2}\}$. For this data set, we cannot distinguish between heterogeneity of survival rates among individuals, and a common pattern of survival for all birds with high losses in the first year after banding. No advantage was found in including all three effects of time, years since banding, and heterogeneity in the survival parameter.

A more realistic model might use reporting rate to represent the effect of time since banding, because of birds still being near the banding location. This is Model $\{r_{t\times a}, S_t\}$ in Table 2, with r_{aj} having $a = 1$ for the first year and $a = 2$ for subsequent years (with no parameter for r_{11}). (Previous authors have used $r_t^\star$ for the first year and r_t for subsequent years.) An alternative with fewer parameters would be a main effects version $\{r_{t+a}, S_t\}$, with

$$\log\left(\frac{r_{ta}}{1 - r_{ta}}\right) = \mu' + \tau_j' + \alpha_a'$$

where the prime denotes parameters for r rather than S.

Table 2 shows that $\{r_{t+a}, S_c\}$ is actually preferred to $\{r_{t+a}, S_t\}$ (test statistic 7.763, 8 d.f., $p = 0.4570$). This simpler $\{r_{t+a}, S_c\}$ model is also very close to the 'best' models. For example, comparing it with $\{r_t, S_c\}$ gives $p = 0.0126$ (test statistic 6.224 on 1 d.f.). This is not quite as useful as the move from $\{r_t, S_c\}$ to $\{r_t, S_{h_2}\}$ tested above ($p = 0.0051$), but it is comparable.

The main effects version r_{t+a} is proving better than the interactive $r_{t\times a}$, both with S_c ($p = 0.0798$) and with S_i ($p = 0.0921$). This could be of interest to modellers even when no heterogeneity is being incorporated: if there is a consistent pattern of recovery rates being higher in the first year since banding, there is a saving of parameters to be made in using the main effects rather than the interactive model.

6 Appraisals

To appraise the models, we ran 100 simulations from each of three scenarios. All three had r_t, and the assumptions for S were S_c, S_{h_2} and $S_{h.\beta}$ respectively. In each case, the input parameters were those estimated for the San Luis male adult mallard data. Each simulated recovery matrix was analysed by the same three models.

Model selection from the simulated data gives an idea of power and test size in this region of the parameter space. With data generated from homogeneous survival rates, S_c, the non-standard likelihood ratio test at a nominal $\alpha = 0.05$ selected S_{h_2} over S_c in 13% of cases, and $S_{h.\beta}$ over S_c in 9% of cases. The true size of each of these tests is higher than the assumed 5%, with S_{h_2} being more likely than $S_{h.\beta}$ to 'detect' heterogeneity when none is there. If the simulated data have S_{h_2} heterogeneity, we found S_{h_2} selected over S_c in 89% of cases, and $S_{h.\beta}$ over S_c in 79% of cases. With data generated from the $S_{h.\beta}$ model, these percentages were 88% and 87% respectively. Both S_{h_2} and $S_{h.\beta}$ are powerful at detecting heterogeneity around these parameter values. If the skewness of the individual survival rates distribution matches that of the fitted beta distribution, the powers are similar. If, however, the skewness does not match, S_{h_2} with its extra parameter is better able to fit the third moment, and this model has more power than $S_{h.\beta}$.

Basic parameter estimates and their standard errors (rounded to three decimal places) are in Table 3. Analysis by the heterogeneous models gives higher standard errors. With true model S_{h_2}, analysis by S_{h_2} shows positive bias in π_1 and S_1 - the maximum likelihood procedure is unwilling to signal a very small group with practically zero survival. Even with data generated from two groups, analysis by that same model shows an inability to retrieve those groups; this seems typical of the use of finite mixtures in capture-recapture (Pledger, 2000) and band recovery models. The average survival in that first year, however, is less biased:

$$\hat{\pi}_1\hat{S}_1 + \hat{\pi}_2\hat{S}_2 = 0.151 \times 0.136 + 0.849 \times 0.6715 = 0.627$$

TABLE 3. Basic parameter estimates and their standard deviations from the simulations. All the models use r_t for the reporting rates

Input model	Analysis model	S	π_1	π_2	S_1	S_2	α	β	$\bar{r}$
S_c	Inputs:	0.638							0.184
S_c	S_c	0.636							0.184
		(0.017)							(0.008)
S_c	S_{h_2}		0.195	0.805	0.419	0.713			0.203
			(0.293)	(0.293)	(0.287)	(0.122)			(0.055)
S_c	$S_{h.\beta}$						29 025	16 396	0.187
							(45 334)	(25 805)	(0.010)
S_{h_2}	Inputs:		0.080	0.920	0.006	0.678			0.188
S_{h_2}	S_c	0.636							0.184
		(0.018)							(0.008)
S_{h_2}	S_{h_2}		0.151	0.849	0.135	0.715			0.204
			(0.132)	(0.132)	(0.179)	(0.082)			(0.067)
S_{h_2}	$S_{h.\beta}$						1 051	593	0.214
							(10 446)	(5 891)	(0.046)
$S_{h.\beta}$	Inputs:						4.450	2.314	0.204
$S_{h.\beta}$	S_c	0.637							0.185
		(0.017)							(0.008)
$S_{h.\beta}$	S_{h_2}		0.290	0.710	0.337	0.789			0.240
			(0.197)	(0.197)	(0.203)	(0.117)			(0.108)
$S_{h.\beta}$	$S_{h.\beta}$						150.491	89.769	0.212
							(1 434.824)	(859.416)	(0.028)

compared with an input value of $0.0797 \times 0.0064 + 0.9203 \times 0.6783 = 0.625$. This is consistent with results in finite mixtures that show that allocation to supposed groups is often unreliable, although estimation of average parameters may be good (Lindsay, 1995). When we use a two-group model, we do not suppose there really are two groups.

Similarly, if $S_{h.\beta}$ (with parameters α and β) is used for both simulation and analysis, the estimate of average survival in the first year is reasonably close:

$$\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = \frac{150.491}{150.491 + 89.769} = 0.626$$

compared with an input of $4.450/(4.450 + 2.314) = 0.658$. This occurs despite the very flat likelihood surface in the α and β directions, as shown by their highly variable and biased estimates.

A derived parameter vector of interest is the average survival rate $\bar{S}$ of a cohort through time, since heterogeneity causes this to rise as birds with lower survival rates are lost from the population. The predicted trend in $\bar{S}$ for Model S_{h_2} starts with $\bar{S}_0$ at time zero. After k years, a proportion $\pi_1 S_1^k$ of group 1 birds has survived, and $\pi_2 S_2^k$ of group 2, giving overall proportions of

$$\pi_{1k} = \frac{\pi_1 S_1^k}{\pi_1 S_1^k + \pi_2 S_2^k} \quad \text{and} \quad \frac{\pi_2 S_2^k}{\pi_1 S_1^k + \pi_2 S_2^k} \quad \text{respectively}$$

The average survival rate after k years is thus

$$\bar{S}_k = \pi_1 S_1^k + \pi_2 S_2^k = \frac{\pi_1 S_1^{k+1} + \pi_2 S_2^{k+1}}{\pi_1 S_1^k + \pi_2 S_2^k} = \frac{E(S^{k+1})}{E(S^k)}$$

say, where the expectation is over the original two-group mixture distribution. Similarly for Model $S_{h.\beta}$,

$$\bar{S}_k = \frac{E(S^{k+1})}{E(S^k)} = \frac{\alpha + k}{\alpha + \beta + k}$$

from the beta distribution moments.

Figures 1, 2 and 3 show these predicted trends in average survival for the three simulation scenarios. Since the trend is being forecast ahead by only a few estimated parameters (π_1, S_1 and S_2 for S_{h_2}, and α and β for $S_{h.\beta}$), they have the potential to become completely wrong over time. Model S_{h_2} will rise until it levels off at the higher survival rate S_2, with an increasing proportion of Group 2 birds surviving, while Model $S_{h.\beta}$ rises towards 1 as its lower survival birds are removed.

If the simulated population actually has constant S (Fig. 1), the heterogeneous models are overly eager to 'detect' heterogeneity, with S_{h_2} performing worse than $S_{h.\beta}$.
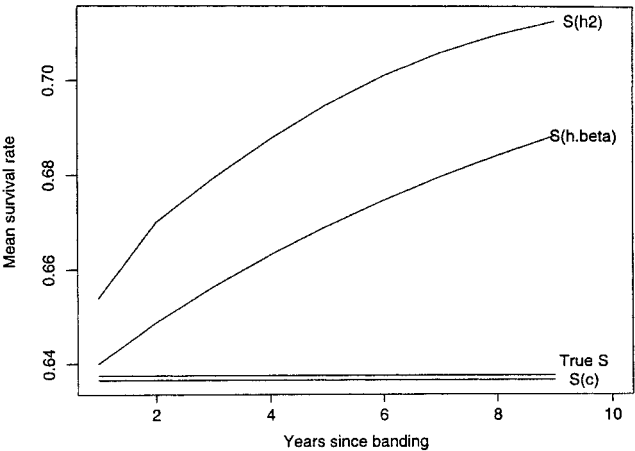
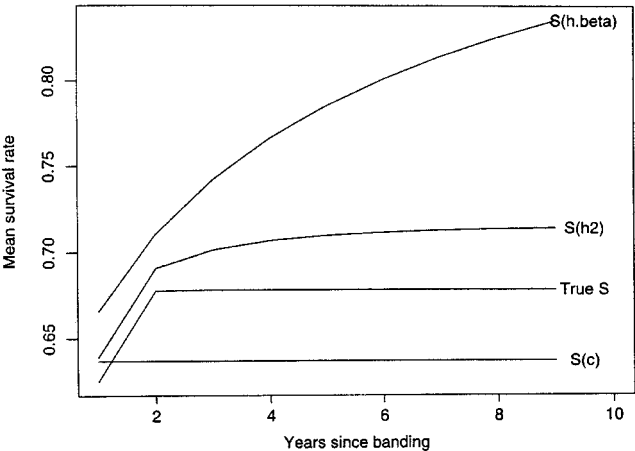FIG. 1. Predicted trend in mean survival rate, data simulated with constant S.

FIG. 2. Predicted trend in mean survival rate, data simulated with S_{h_2}.
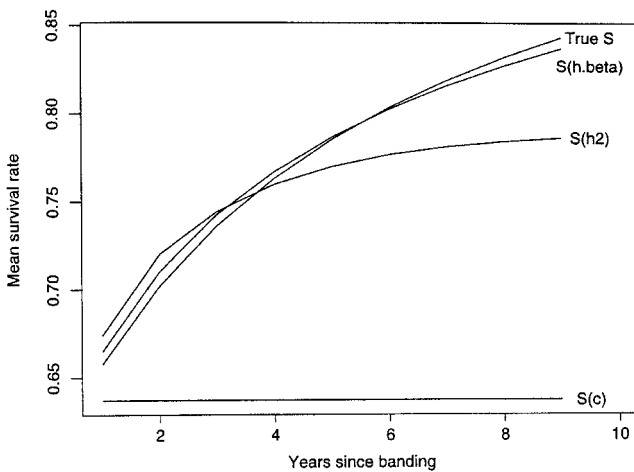
FIG. 3. Predicted trend in mean survival rate, data simulated with $S_{h.\beta}$.

With S_{h_2} simulations (Fig. 2), the S_{h_2} analysis leads to a curve that rises too high, because of its reluctance to allow a very small group with practically zero survival, while the beta model rises far too high as it goes towards 1.

Figure 3 shows the $S_{h.\beta}$ simulations. The $S_{h.\beta}$ analysis is remarkably close to the input population despite the unreliable α and β estimates, while S_{h_2} is lower.

The relative performance of the two-group and the beta models for heterogeneity will depend on the skewness and higher moments of the distribution of S. If these happen to match the skewness and higher moments of the two-parameter beta distribution, the beta analysis is likely to perform better. If, however, the skewness of the true distribution does not match the beta, the extra parameter of the two-group mixture is likely to help. In addition, more than two groups may be selected, if there are enough data to signal an improved fit.

7 Discussion

These methods provide a way to deal with 'hidden' heterogeneity in the datasets. Explicit heterogeneity that is related to variables known only at the time of recovery (e.g. wintering areas in migrating birds) could be handled by post-stratification (Schwarz *et al.*, 1988). Heterogeneity related to variables known at the time of banding (e.g. gender) could be handled by the method of Lebreton *et al.* (1992). Consequently, a two-group model could be used to assess if a covariate-based model has sufficient complexity to capture most of the heterogeneity in survival and recovery rates that may be present.

If overdispersion is present in the data, with an overdispersion parameter above say 3 or 4, one possible cause is parameter heterogeneity, with individuals having unique parameters rather than the same one (Burnham & Anderson, 1998). This indicates that heterogeneous models could be fitted, and then compared with models using an overdispersion parameter for their ability to correct the standard error estimates.

These models also have immediate usefulness to assess acute handling mortality or immediate tag loss. For example, the San Luis dataset indicated that a small group of animals (about 8%) had essentially a zero survival rate. This could be

indicative of either handling mortality or band loss. The act of tagging is very stressful for deep water fish. These models could also be used to cross validate estimates of initial tag shedding based on experiments from a mixture of single and double tagged fish and multiple tag types (Barrowman & Myers, 1996). However, Arnason & Mills (1981) investigated the effects of continual tag loss and showed that continual tag loss is confounded with mortality and leads to a downward bias in the estimates. Consequently, these models would not be able to detect this type of tag loss if loss operates independently of mortality—but will be able to detect tag loss if loss is related to area of banding, type of band used, person who did the banding, etc.

The method could also be extended to Cormack-Jolly-Seber experiments. For example, these models could easily account for transients, animals that move in and out of the study population quickly. These would appear from the model fitting in a similar fashion to the San Luis dataset, as a fraction of the animals with very low apparent survival rate. A slight modification of our approach gives the model of Pradel *et al.* (1997), in which there were two groups defined, transients and residents. Those with more than one capture were assigned to the resident group, while those caught once were modelled as a two-group mixture, some (the transients) with zero survival rate, and the rest with survival rate matching that of the known residents. If there really are two such groups, the two slightly different approaches to the modelling should give similar results.

These finite mixture models offer considerable scope for model building, and should find applications in many areas where heterogeneity is present.

Acknowledgements

REFERENCES

ARNASON, A. N. & MILLS, K. H. (1981) Bias and loss of precision due to tag loss in Jolly-Seber estimates for mark-recapture experiments, *Canadian Journal of Fisheries and Aquatic Science*, 38, pp. 1077-1095.

BARROWMAN, N. J. & MYERS, R. A. (1996) Estimating tag-shedding rates for experiments with multiple tag types, *Biometrics*, 52, pp. 1410-1416.

BROWNIE, C., ANDERSON, D. R., BURNHAM, K. P. & ROBSON, D. S. (1985) *Statistical Inference from Band-Recovery Data—A Handbook*, 2nd edn (Washington DC, Fish and Wildlife Service Resource Publication 156).

BURNHAM, K. P. & ANDERSON, D. R. (1998) *Model Selection and Inference* (New York, Springer-Verlag).

BURNHAM, K. P. & REXSTAD, E. A. (1993) Modeling heterogeneity in survival rates of banded waterfowl, *Biometrics*, 49, pp. 1194-1208.

CATCHPOLE, E. A., FREEMAN, S. N. & MORGAN, B. J. T. (1995) Modelling age variation in survival and reporting rates for recovery models, *Journal of Applied Statistics*, 22, pp. 597-609.

DAVIES, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, pp. 247-254.

DAVIES, R. B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 74, pp. 33-43.

LEBRETON, J.-D., BURNHAM, K. P., CLOBERT, J. & ANDERSON, D. R. (1992) Modelling survival and testing biological hypotheses using marked animals. A unified approach with case studies, *Ecological Monographs*, 62, pp. 67-118.

LINDSAY, B. G. (1995) *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional

Conference Series in Probability and Statistics 5. Institute of Mathematical Statistics, Hayward, California.

LINDSAY, B. G. & ROEDER, K. (1992) Residual diagnostics for mixture models, *Journal of the American Statistical Association*, 87, pp. 785-794.

NICHOLS, J. D., STOKES, S. L., HINES, J. E. & CONROY, M. J. (1982) Additional comments on the assumption of homogeneous survival rates in modern bird banding estimation models, *Journal of Wildlife Management*, 46, pp. 953-962.

NORRIS, J. L. & POLLOCK, K. H. (1995) A capture-recapture model with heterogeneity and behavioural response, *Environmental and Ecological Statistics*, 2, pp. 305-313.

NORRIS, J. L. & POLLOCK, K. H. (1996) Nonparametric MLE under two closed capture-recapture models with heterogeneity, *Biometrics*, 52, pp. 639-649.

PLEDGER, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics*, 56, pp. 434-442.

POLLOCK, K. H. & RAVELING, D. G. (1982) Assumptions of modern band-recovery models, with emphasis on heterogeneous survival rates, *Journal of Wildlife Management*, 46, pp. 88-98.

PRADEL, R., HINES, J. E., LEBRETON, J.-D. & NICHOLS, J. D. (1997) Capture-recapture survival models taking account of transients, *Biometrics*, 53, pp. 60-72.

RICHARDSON, S. & GREEN, P. J. (1997) On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society, Series B, Methodological*, 59, pp. 731-758.

SELF, S. G. & LIANG, K.-Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association*, 82, pp. 605-610.

SCHWARZ, C. J., BURNHAM, K. P. & ARNASON, A. N. (1988) Post-release stratification in band-recovery models, *Biometrics*, 44, pp. 765-785.